BOGDAN HARHĂTĂ*

COLLOCATIONS AND DICTIONARIES: A CASE-STUDY APPLIED TO ROMANIAN[1]

It is common apprehension that nowadays dictionaries are becoming ever more widespread and necessary in an array of fields varying from learning/teaching to highly specialized translations. Less is known, on the contrary, of the amount of work and knowledge a dictionary implies. Often, it may happen that so overwhelming can the effort to compile a dictionary be, that the result of such effort would lag behind the newer advances made in theoretical lexicology. Such is the case of Romanian lexicography that has been evolving, over a period stretching well above a century, round Academy's dictionary (namely the DA/DLR).

Either due to the need for maintaining the identity of the dictionary, or to the more trivial reason that Romanian lexicographers have always been too few, and that they just didn't have time and resources to elaborate more sophisticated theoretical approaches, as well, we are now facing the rather awkward situation that some noteworthy differences may be acknowledged between the theoretical fundamentals of what is reputed to be the single most significant lexicographical work of the Romanian language and newer (and not so newer) distinctions widely accepted by mainstream lexicologists.

One such example is the so-called *collocation*, which is yet to find its place in Romanian lexicology and lexicography.

Establishing what exactly a collocation is, and whether the concept itself may find application in Romanian[2] is, for a first step, in order.

The collocations are but one type of restrictive combinations of words. Several definitions of it are available, yet general consensus among linguists is still

* **Bogdan Harhătă**: Scientific Researcher; Lexicology-Lexicography, Classical Studies.

[1] This paper comes as a result of the work at Romanian Academy's *DLR*, and therefore is aimed to cover the practitioner's interest, rather than a broader theoretical point of view.

[2] The Dicț. șt., p. 114 describes the peripheric position of the concept of collocation in Romanian linguistics: „*Termenul este utilizat mai ales în semantica anglo-americană şi cu totul izolat în lingvistica românească*. [The term is frequently used in Anglo-American semantics and sparsely in Romanian linguistics]". Although such claim seems to rule out the very need of introducing the concept in Romanian lexicology, it is nothing else but a mere result of the fact that the writings in Romanian lexicology are few, and, for the most part, heavily influenced by Academy's DLR and DEX.

DACOROMANIA, serie nouă, XVII, 2012, nr. 2, p. 147–153

to be met[3]. We chose to illustrate here three such definitions, starting from a „broader" one, which comprises any frequently encountered combination of words, to a very narrow one, which sees collocations as custom-established word combinations, which convey by preference a determinate meaning.

The broader definition is a rather statistical one: it regards collocations as high-frequency co-occurrences of two or several words in a given language[4]. Such statistic-biased characterization may prove interesting as a starting point, but it fails to describe in a more appropriate manner the lexical structure of a given language, since, according to such definitions, one may rightly argue that examples like *am* + *mâncat*, *voi* + *dormi* (AUX. + verb)[5] or *să* + *fii* (CONJ. + verb)[6] or *foarte* + *bun/bine* (ADV. + ADJ./ADV.)[7] are collocations, which, at least in the meaning we intend to give the term here, they are not.

Further more accurate defining was provided by Igor Mel'cuk and Leo Wanner[8]. According to such a point of view, a collocation would consist of the co-occurrence of two/several words subject to one or several restrictive rules. Albeit being by far more precise than other authors[9], since it does stress out that words being more frequently than usual encountered together do so on account of the fact that they are subject to rule of restriction on combination, this definition omits to specify which is the actual restriction type involved.

A third, more accurate, definition of the collocation suggests that it consists of a word combination, which is subject to a lexical restriction, and therefore, if a given word (the collocated term) is to convey a determined meaning, then such choice is governed/required by a second term (the collocator), which is described by such meaning. A good example would be the wording Rom. *lacrimi amare* [lit. bitter tears] used for expressing the hopeless pain. *Lacrimi* [tears], which is here the

[3] Like many other concepts that permeated terminology in the second half of the last century, different schools of thought still argue on the exact meaning of the term. Of the variety of possible definitions, at least five are generally accepted among specialist. Since the aim of this paper is not that of bringing into discussion the theoretical basis of the concept itself, we followed here, with the due revisions, the definitions proposed by Ježek 2005, pp. 177–180, who reduced the number, for didactic purposes, to three. It is noteworthy that Italian lexicologists seem to have reached consensus upon establishing a technically satisfactory definition applicable to Italian. See, also: Beccaria Diz., Casadei 2003, de Mauro 2005 etc.

[4] That is the definition provided by Benson *et al.* 1986, p. IX. The actual wording by the authors is: *recurrent* [word] *combinations*.

[5] *am mâncat* = *am* (I have – auxiliary) + *mâncat* (Past Participle of *mânca* = eat) / *voi dormi* = *voi* (I will) + *dormi* (Infinitive of *dormi* = sleep).

[6] In Romanian the conjunction *să* function as morpheme of the Subjonctive.

[7] The comparison degrees in adjectives and adverbs Romanian include an adverb *mai* [more] or *foarte* [very] + ADJ./ADV.

[8] Mel'cuk, Wanner, p. 325; the actual wording of the authors is *restricted lexical co-occurrence*.

[9] Beccaria Diz., pp. 154–155 describes the way the term collocation is used in the anglo-saxon literature in the following words: „Nella linguistica inglese [la collocazione] è stata usata spesso come un iperonimo di qualsiasi combinazione di parole, dalle strutture verbo + preposizione obbligatoriamente richieste da certi verbi, alle mere solidarietà lessicali, alle frasi idiomatiche, ai proverbi e alle formule fisse del tipo *Come stai?, Buon giorno, Pronto, Chi parla?*".

collocator, prefers to combine with the adjective *amar* [bitter] (collocated term), rather than with other adjectives that may ideally replace it (like, for example *dureroase* [painful], or *deznădăjduite* [hopeless]). A different kind of example involves a given noun that requires a certain verb. We have thus phrases like *a întocmi un document* [draw up <to> a document], where, in order to convey the making of it, *documentul* [the document] (collocator) rather „demands" a specific verb, i.e.: *a întocmi* [lit. put together <to>] (collocated term), instead of other semantically compatible verbs, i.e. *a scrie* [write <to>], or *a redacta* [compose, compile <to>].

Ježek argues that, as complex this definition might be, it still doesn't answer in a satisfactory manner to the needs of specificity of the term, since it implies that semantic and lexical solidarity, on one hand (*a îmbrăca* and *haină*), and the true collocation are one and the same[10]. Further distinction is, indeed, in order: one needs to asses the differences between cases of lexical solidarity[11] and collocations. Contrasting sets of examples may prove useful here:

In the case of semantic and lexical solidarities,

when taken alone

a parca [park <to>]	implies	*maşină/automobil* [car];
a îmbrăca [dress <to>]	implies	*îmbrăcăminte, haine* [clothing];
acvilin [aquiline]	implies	*nas* [nose];
blond [blond]	implies	*păr* [hair];
câine [dog]	implies	*a lătra* [bark <to>].

In the case of true collocations,

when taken alone

a întocmi [write <to>; draw up <to>]	doesn't necessarily imply	*document* [document];
a lansa [launch <to>]	doesn't necessarily imply	*mesaj* [message];
amar [bitter]	doesn't necessarily imply	*lacrimă* [tear];

[10] Ježek 2005, p. 178.

[11] The concept was presumably introduced by Coseriu – see. Coseriu 1969, passim – as a means to explain the mechanisms activated when meaning relations within given word pairs are observed. The English reader might be more familiar with the wording «selectional restriction» forged by Noam Chomsky in 1965 on the basis of the notion of «selection» already put into circulation by Louis Hjelmslev in 1961. In fact, despite the novelty of the interpretation, the observation that meaning relations that pre-condition the syntagmatic level of speech/text go back to Walter Porzig, in 1934, who is often cited as the first to have them systematically identified. The *ad-litteram* translation, indeed, of the term used by Porzig, i.e.: *Bedeutungsbeziehung*, gives nothing else but *meaning connexion*, which explains rather well, in itself, the possible signification of the lexical solidarity.

All the examples above show a lexical restriction (i.e. a preference for entering combinations with a specific given word). In which way is the nature of such restriction different in the two cases? If, in case of the semantic solidarities, there is always a reciprocal meaning implication, which remains active for the collocated term even when used alone (*îmbrăca* [dress/wear <to>] necessarily implies *îmbrăcăminte* [some piece of clothing], while the *câine* [dog] intrinsically *latră* [barks]), when true collocations are involved, the reciprocal meaning implication is given by the combination alone, and is absent when one of the terms is missing (*a întocmi* [lit. put together <to>] doesn't necessarily imply a *document* [document, act]). It is, also, noteworthy noticing that in Romanian (albeit a historically motivated different distribution throughout vocabulary, the situation is present in the rest of the Romance languages), words like *a întocmi, a lansa, amar* are polysemic, while words like *a îmbrăca, acvilin* etc. are monosemic.

The field of application that first comes to mind when trying to find the usefulness of treating collocations as a somewhat separate lexical entry is that of the contrastive language learning, since different languages operate quite differently when they need to choose the collocated terms, in other words different languages intimately organize their lexical material in different ways. Why contrastive language learning? Because, beyond language-inherent lexical choices, some resemblances between two different languages leave the door open to interpretation by way of recognition, e.g. should a foreign learner of Romanian be familiar with the raw meaning of the word *amar*, he/she could be able to figure out the meaning of the sequence in original *a plânge cu lacrimi amare*, yet some need further explanation that cannot be reconstructed by way of operating selections in the structure of senses of the words in a collocation. One example I stumbled myself upon while learning Italian was the wording *pioggia battente* [heavy rain][12], which is called so on account that the falling rain produces a considerable amount of noise, if heavy enough. If this latter example could bear some logical explanation, tough remote and intricate, further examples have etymological justifications so far in time that nowadays appear unexplainable and arbitrary; e.g.: the Romanian wording for a dark night is *noapte adâncă* [lit. *deep night*], the Italian one is *notte fonda*, and in no circumstances *notte profonda*, which, although possible, is stylistically marked and has a different meaning.

The usefulness of defining in a standardized manner collocations and in assigning them a specific treatment in dictionaries brings us to the point we started from. As promising as contrastive language learning, translation etc. may be, we should not forget that bilingual dictionaries are always based on standard monolingual ones. Lexicographers, therefore, if not for the sake of acknowledging the advances made by more theoretical fields of linguistics alone, need to adjust their methods to the actual purpose of the nowadays used daily, either in a printed

[12] The example appears, also, in Ježek 2005, p. 179.

form or online, object, which is the dictionary, i.e. that of a working instrument. The situation is particularly complex in the case of Romanian, on one hand due to the fact that the field of theoretical lexicology is still missing some of its essential instruments and that Romanian lexicologists are traditionally trained to be first lexicographers, and on the other because the efforts of the Romanian lexicographers were aimed at compiling the *Thesaurus* of the Romanian language, namely Academy's DLR. Technical reasons stay behind the fact that collocations haven't yet come to find a place in nowadays Romanian dictionaries, among these, the most significant one being that the Academy's DLR (and the other dictionaries that evolved round it) was due to maintain its structural unity over a period longer than one hundred years. There is, also, no need to add that more practical fields of linguistics tend (like is the case in any other science) to „resist" last-hour innovation, and are more conservative, nor that the most part of the established lexicographic traditions are still seeing words as units in exclusively paradigmatic systems, which is particularly convenient when one needs to arrange words and their meanings in a *systematic* written form.

The Romanian Academy's DLR gives attention to the syntagmatic behaviour of words, since it registers the so-called *expresii* [verbal idiomatic phrases], *locuţiuni* [adverbial, prepositional etc. phrases], and the compound words which include the entry title. None the less, the collocations fail to be specifically indicated, and are listed among other quotations under the due meaning of the word. We chose to exemplify here with the word *umbră* [shadow], for two reasons: the letter U was published in the recent years (the volume was issued in 2002), and the word itself is an old one, with abundant occurrences in the literary sources. The word stretches over eight pages of the tome, from page 109 to page 116 and provides quotations from sources as old as 1643[13] until nowadays. One would expect, statistically, to encounter quite frequently in the several hundreds quotations the collocation that is standard in Romanian for saying that the shadow is deep, i.e.: *umbră deasă* [lit. thick shadow][14]. Oddly, the (unmarked) collocation appears only once, in the very first (and oldest) source, namely in the Homilies of Varlaam: *La mijlocul calei se află un copac frumos şi cu umbră deasă* [Half the way through, there is a beautiful tree that casts a deep shadow]. Equally oddly, the antonymic collocation, i.e.: *umbră rară* [lit. thin shadow] appears, also, only once, and in a quotation from a text from the year 1673[15]. The conclusions the unadvised reader may draw are forthright:

[13] Varlaam Mitropolitul, *Carte românească de învăţătură, Dumenecele preste an şi la praznice împărăteşti şi la yvenţi mari. Cu dzisa şi cu toată cheltuiala lui Vasilie voivodul şi domnul Ţărâi Moldovei din multe scripturi tălmăcită din limba slovenească pre limba romeniască de Varlaam Mitropolitul de Ţara Moldovei, în tiparul domnesc, în Mănăstirea a Trei S[feti]teli în Iaşi de la Hs. 1643* [The Homilies of Varlaam, Metropolitan Bishop of Moldavia].

[14] Curiously enough, other Romance languages prefer the same word combination, cf. it. *ombra fitta*, fr. *ombre epaisse*.

[15] Dosoftei, *Psaltirea în versuri 1693* [The Book of Psalms by the Metropolitan Bishop Dosoftei].

– the collocation is old and no longer in use;

– the collocation is rare enough, and its frequency of use in language is rather low (one should, also, mention, that there is no quotation from folklore sources[16]).

Instead, to any educated Romanian speaker's/reader's ear, the wording *umbră deasă/rară* sounds natural, because the native speaker's instinct (competence) tells so; that is a wording frequently encountered in both cult literature and folklore, and alive in the spoken language. Much emphasis was placed on the importance of making collocations visible in bilingual dictionaries, we reckon one mustn't forget that monolingual dictionaries, also, should register, preserve and, first of all, cultivate such specific characteristics of a given language.

Does collocation need to be acknowledged in Romanian linguistics and, consequently, given a specific place in Romanian dictionaries? If the answer form nowadays lexicographers will be affirmative, the advantages would outnumber the technical difficulties. Among these one may count: catching up with lexicographic traditions in the major Romance languages[17], easier and better organized lexical research, as well as more practical goals like easier teaching/learning Romanian, or facilitating adequate translation from/into Romanian.

REFERENCES

Aprile 2008 = Marcello Aprile, *Dalle parole ai dizionari*, Bologna, il Mulino, 2008.

Beccaria Diz. = Gian Luigi Beccaria (ed.), *Dizionario di linguistica e di filologia, metrica, retorica*, Torino, Einaudi, 2004.

Benson *et al.* 1986 = Morton Benson, Evelyn Benson, Robert Ilson, *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*, Amsterdam – Philadelphia, John Benjamins, 1986.

Casadei 2003 = Federica Casadei, *Lessico e semantica*, Roma, Carocci, 2003.

Comrie 1981 = Bernard Comrie, *Language Universals and Linguistic Typology*, Oxford, Blackwell, 1981.

Corda, Marello = Alessandra Corda, Carla Marello, *Imparare e insegnare il lessico*, Torino, Paravia, 1999.

Coseriu 1969 = Eugenio Coseriu, *Solidarietà lessicali*, in idem, *Teoria del linguaggio e linguistica generale. Sette studi*, Roma – Bari, Laterza, 1971, pp. 303–316.

Cruse 2002–2005 = Alan D. Cruse, Franz Hundsnurcher, Michael Job, Peter Rolf Lutzeier (eds.), *Lexicologie/Lexicology*, 1. Halbband 202, 2. Halbband 2005, Berlin – New York, Mouton – de Gruyter, 2002–2005.

de Mauro 2000 = Tullio de Mauro, *Dizionario della lingua italiana*, Torino, Paravia, 2000.

de Mauro 2005 = Tullio De Mauro, *La fabbrica delle parole: il lessico e problemi di lessicologia*, Torino, UTET, 2005.

Dicț. șt. = Angela Bidu-Vrănceanu, Cristiana Călărașu, Liliana Ionescu-Ruxăndoiu, Mihaela Mancaș, Gabriela Pană Dindelegan, *Dicționar de ştiințe ale limbii*, Bucureşti, Nemira, 2001.

DLR = *Dicționarul limbii române* (serie nouă), București, Editura Academiei R.P.R./R.S.R./Române, 1965 și urm.

Firth 1951 = John R. Firth, *Modes of Meaning*, in idem, *Papers in linguistics 1934–1951*, London, Oxford University Press, 1957, pp. 190–215.

[16] The wordings *umbră deasă/rară* are standard formulations in folklore.

[17] Mostly with the Italian one, which has the advantage of a particularly good integration of the theoretical lexicology with lexicography, as well as of a remarkable unity around defining functional concepts.

Graffi, Scalise = Giorgio Graffi, Sergio Scalise, *Le lingue e il linguaggio: introduzione alla linguistica*, Bologna, il Mulino, 2002.
Ježek 2005 = Elisabetta Ježek, *Lessico. Classi di parole, strutture, combinazioni*, Bologna, il Mulino, 2005.
Lo Cascio = Vincenzo Lo Cascio, *Semantica lessicale e criteri di collocazione nei dizionari bilingui a stampa ed elettronici*, in Tullio de Mauro e Vincenzo Lo Cascio (eds.), *Lessico e grammatica*, Roma, Bulzoni, 1997, pp. 63–88.
Lyons 1968 = Jeremy Lyons, *Introduction to Theoretical Linguistics*, Cambridge, Cambridge University Press, 1968.
Marello 1996 = Carla Marello, *Le parole dell'italiano*, Bologna, Zanichelli, 1996.
Mel'cuk, Wanner = Igor Mel'cuk, Leo Wanner, *Towards and Efficient Representation of Restricted Lexical Co-occurrence*, in Willy Martin *et al*. (eds.), *Proceedings of the 6th EURALEX International Congress on Lexicography*, Amsterdam, John Benjamins, 1994, pp. 325–338.
Urzě = Francesco Urzě, *Dizionario delle Combinazioni Lessicali*, Luxembourg, Convivium, 2009.
Voghera 1994 = Miriam Voghera, *Lessemi complessi: percorsi di lessicalizzazione a confronto*, in „Lingua e stile", XXIX, 1994, pp. 185–214.

Abstract

This paper aims to bring into discussion the need to introduce in Romanian lexicology/ lexicography the concept of collocation. The paper briefly provides a convenient definition of the term, applicable to the particular characteristics of Romanian and other Romance languages, and argues that placing collocations in a visible manner within Romanian dictionaries could prove useful to both theoretical approaches, and practical ones, like language learning and translation.

Keywords: *collocation, lexicology, lexicography, Romanian language, dictionary, word combinations, lexical restriction, restrictive word combination.*

Institutul de Lingvistică şi Istorie Literară „Sextil Puşcariu"
400165 Cluj-Napoca, str. Emil Racoviţă, 21
România
bogdan_harhata@inst-puscariu.ro